



Д.В. Журавлёв, В.С. Смолин

**Нейросетевая революция
искусственного интеллекта и
варианты её развития**

Рекомендуемая форма библиографической ссылки

Журавлёв Д.В., Смолин В.С. Нейросетевая революция искусственного интеллекта и варианты её развития // Проектирование будущего. Проблемы цифровой реальности: труды 6-й Международной конференции (2-3 февраля 2023 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2023. — С. 223-244. — <https://keldysh.ru/future/2023/16.pdf>
<https://doi.org/10.20948/future-2023-16>

Размещено также [видео выступления](#)

Нейросетевая революция искусственного интеллекта и варианты её развития

Д.В. Журавлёв¹, В.С. Смолин²

¹ ООО ЦИФРОМЕД, Служба управления данными

² Институт прикладной математики им. М.В. Келдыша РАН

Аннотация. Развитие цивилизации, которое обеспечивает прогресс наук и технологий, привело к созданию теорий алгоритмов и управления и достижению возможностей создания «умных» машин, в первую очередь, на основе информационно-вычислительных технологий. Но аналогично тому, как в химии до первой половины XIX в. считалось, что для получения органических веществ необходима «жизненная сила», так и сейчас распространено мнение, что все современные успехи искусственного интеллекта – это имитация умственной деятельности, которая «по настоящему» может происходить только в мозгу живого человека. Для обоснования этого мнения указывается, что у созданных сетей формальных нейронов и алгоритмы работы заметно отличаются от функционирования живых нейросетей и свойства процессов их обучения тоже кардинально другие. Да, это так, но современные успехи развития «узкого» искусственного интеллекта дают основу для понимания имеющихся нерешённых проблем и позволяют искать пути их преодоления. Главный шаг состоит в отказе от имитации способов решения задач на основе имеющихся у человека знаний и переходе к получению и накоплению новых знаний путём взаимодействия со средой. Но это не значит, что цивилизационные знания, накопленные человечеством, следует отбросить. Наоборот, именно имеющиеся знания должны служить базой для получения новых.

Ключевые слова: формальные нейроны, сильный искусственный интеллект (СИИ), цивилизация, конкуренция, декомпозиция, знания

The neural network revolution of artificial intelligence and it's development options

D.V. Zhuravlev¹, V.S. Smolin²

¹ CIFROMED LLC, Data Management Service

RAS Keldysh Institute of Applied Mathematics

Abstract. The civilization development, which ensures the science and technology progress, has led to the theories of algorithms and control creation

and the achievement of technological capabilities to create “smart” machines, primarily based on information and computing technologies. But, just as in chemistry until the first half of the 19th century there was no doubt that “life force” was needed to obtain organic substances, so it is now widely believed that all modern successes in artificial intelligence are only an imitation of mental activity, which “for real” can occur only in a living person brain. To substantiate this opinion, it is indicated that the created networks of formal neurons not only have work algorithms that differ markedly from the living neural networks functioning, but also their learning processes properties are also radically different. Yes, this is true, but modern successes in the development of “narrow” artificial intelligence provide a basis for understanding the existing unresolved problems and allow us to look for ways to overcome them. The main step is to abandon the imitation of problem solving methods based on the knowledge that a person has and the transition to the acquisition and accumulation of new knowledge through interaction with the environment. But this does not mean that the civilizational knowledge accumulated by mankind should be discarded. On the contrary, it is the existing knowledge that should serve as the basis for obtaining new ones.

Keywords: formal neurons, strong artificial general intelligence (AGI), civilization, competition, decomposition, knowledge

1. Введение

Развитие в начале 2-й четверти XX в. теории алгоритмов на основе «универсальной машины Тьюринга» [1] показало возможность создания устройств, способных (при наличии достаточных ресурсов) реализовать любой алгоритм. В 1950-е гг. компьютеры, работающие по схеме фон Неймана, реализующей идеи «машины Тьюринга», получили распространение и нашлись энтузиасты, желающие исследовать вопрос возможности описания «интеллектуальной» деятельности человека формальными алгоритмами. Важной вехой в активизации работ по данному направлению стал Дартмутский семинар 1956 г. (Гановер, США). Целью семинара, длившегося 2 месяца, было рассмотрение вопросов: можно ли моделировать рассуждения, интеллект и творческие процессы с помощью вычислительных машин. Хотя эйфория, связанная с надеждой, что данные вопросы на семинаре удастся решить, не оправдалась, практически были сформированы основные направления развития и сам термин искусственного интеллекта (ИИ).

2. Нисходящий и восходящий подходы к развитию ИИ

Наиболее прямым и коротким путём к созданию ИИ изначально представлялась программная реализация алгоритмов на основе интерпретации человеческих знаний о правилах работы с известными понятиями и закономерностями. Данное направление получило название «нисходящий подход», т.к. для его развития необходимо брать высокоуровневые абстракции и искать пути их применения к низкоуровневым «сырым» данным. И первые 50 лет практического развития ИИ «нисходящий подход», основанный

на разработке эвристических алгоритмов применения «высоких» идей к прикладным задачам, давал наиболее заметные практические результаты.

Альтернативой является «восходящий подход», предполагающий, что на основе обработки низкоуровневых «сырых» примеров соответствия между входом X и выходом Y можно выявить «высокие» идеи их взаимосвязи. Ведь человечеству, используя нервные системы людей, удалось такие идеи сформировать. «Нейросетевые» алгоритмы используют различные архитектуры, состоящие в основном из формальных нейронов.

Схема формального нейрона, как логического преобразователя бинарных сигналов, была опубликована в 1943 г. У. Маккалоком и У. Питтсом [3], но сейчас используется её расширение на случай аналоговых (входных и выходных) сигналов, предложенное в 1960 г. Уидроу и Хоффом [4]. Такие аналоговые преобразователи сигналов лучше подходят для обработки данных о непрерывном с точки зрения сенсорного восприятия мире.

Системы, весьма косвенным путём (через реализацию длительного обучения сетей формальных нейронов) достигающие решения «интеллектуальных» задач, стали давать лучшие, чем у эвристических алгоритмов, результаты только в последние 10-12 лет. Но и в годы доминирования эвристических алгоритмов большой интерес околонуточной общественности вызывали даже незначительные успехи нейросетевых моделей.

Поскольку разумные (для прошлого века) доводы о преимуществах «нисходящего подхода» только увеличивали внимание к нейросетям, сторонники эвристических алгоритмов ещё в 1960-е гг. присвоили нейросетевикам кличку “Scruffies” (неряшливые, нечёткие). При этом они себя называли “Neats” (клёвые, чёткие). В русском научном сообществе пошли дальше: кличку для нисходящего подхода перевели как «чистый», для восходящего – «грязный». И сторонники эвристических алгоритмов зачастую до сих пор используют эту терминологию, не понимая, что это выдаёт их нежелание воспринять нейросетевые подходы.

3. Имитация или создание AGI?

Тем не менее, нейросетевая революция в области ИИ состоит в том, что нейросетевые подходы заняли в ней центральное место. Но степень, до которой нейросетевые подходы могут позволить реализовать различные аспекты интеллектуальной деятельности, остаётся темой обсуждения. Можно спорить, что является более, а что менее важным для реализации интеллектуальной деятельности, называть решение сложных задач, понимание, интуицию, эмоции, формирование целей и многое другое. **Центральным является вопрос:** это всегда будет имитация интеллектуальной деятельности или возможно создать системы или агентов ИИ, действительно обладающих интеллектом?

Ещё в обосновании для проведения Дартмутского семинара 1956 г. Дж. Маккарти писал: «любое свойство интеллекта может быть столь точно

описано, что машина сможет его имитировать» [5]. Если тогда ни описать, ни симитировать практически ничего не удалось, то сейчас всё больше свойств интеллекта всё более успешно имитируются. Если удастся симитировать все известные свойства интеллекта, будет ли это всё ещё имитацией или уже построением интеллекта хотя бы человеческого уровня?

4. Материально ли сознание?

Тут в значительной степени ключевым является вопрос терминологии. Калькулятор имитирует или выполняет вычисления?! Вроде бы выполняет, но не как человек, не осознавая, что он делает. А нужно ли для правильных вычислений осознавать процесс? И что значит осознавать? Это важный вопрос, поскольку ИИ реализуется на вычислительной технике, которая ничего, кроме вычислений, делать не умеет. Если все свойства интеллекта можно свести к вычислениям, то их можно не имитировать, а выполнять. А если необходимо сознание – надо не только понять зачем, но и как реализовать сознание с помощью вычислений.

Многие из нас воспитаны в духе материализма: материя первична, сознание – вторично. Поскольку есть ещё и идеалисты, придерживающиеся противоположного мнения, вопрос о соотношении материи и сознания часто называют главным вопросом философии. Этот вопрос не имеет теоретического решения, потому что первичность материи или сознания просто постулируется сторонниками каждого течения.

Существование материи отрицать сложно, солипсизм, то есть признание только собственного индивидуального сознания в качестве единственной и несомненной реальности, принято рассматривать как крайнюю, явно не соответствующую реальности форму субъективного идеализма. А вот вопрос соотношения сознания и материи, является темой многовековых философских споров.

При постулировании первичности сознания, его носителем должна быть некоторая нематериальная сущность, «душа», которая, по представлениям идеалистов, может жить отдельно от тела, в «идеальном» мире, в том числе и после смерти тела субъекта. Объективный, диалектический идеализм признаёт существование и материи, и сознания, причём у каждого субъекта – своего, отдельного от других. Поскольку научных, воспроизводимых данных об отдельном от тела носителе сознания пока не предъявлено, это даёт широкое поле для развития различных идеалистических философских течений на основе постулата о первичности сознания.

При строительстве заборов и охране кукурузы природа сознания совсем не важна, отсутствие научных данных о возможности создания нематериального сознания никак не отражается на выполняемой деятельности. ***Но для построения сильного искусственного интеллекта (СИИ)*** (который, вероятно, должен сознанием обладать) вопрос о природе сознания перестаёт быть риторическим.

Согласно постулату о первичности материи сознание может базироваться на результатах реализованных в ходе эволюции жизни процессов самоорганизации. То есть сознание – это свойство высокоорганизованной материи, и современный взгляд состоит в том, что это свойство реализуется в процессе высшей нервной деятельности в мозге животных и человека.

Но и диалектический материализм вполне позволяет предположить, что сознание основано на некоторой вполне материальной, но пока не доступной для нашего наблюдения сущности. Такими сущностями в конце XIX в. были теплород и эфир, а сейчас – тёмные материя и энергия. Надежды найти явные (а не косвенные, которые, как в случае с теплородом и эфиром, оказалось возможным объяснить и без привлечения этих понятий) подтверждения тёмных материи и энергии для теории развития Вселенной (и физики в целом) сохраняются до сих пор. И нельзя исключить, что эти надежды оправдаются, как случилось с электромагнитными и квантовыми представлениями.

Примерно такая же ситуация с материалистическими представлениями о сознании. Материалистическая позиция исследователя не мешает ему строить свои теории о природе сознания на основе пока не изученных квантовых явлениях или той же тёмной материи. Поскольку данная область недостаточно изучена, мы не можем судить о её перспективности. Но и использовать результаты, которые только есть надежда получить, для построения СИИ пока невозможно. Все практические результаты в области ИИ сегодня получают на основе использования нейросетевых алгоритмов, которые алгоритмически моделируют электрохимическую активность нервной системы (НС) животных и человека.

Большинство исследователей согласно с тем, что а) учёные должны создать теорию СИИ; б) технологи – разработать способы физической реализации выдвинутых учёными идей; в) инженеры – спроектировать и наладить выпуск технических устройств – агентов СИИ.

В процессе развития цивилизации человечеству удалось достичь значительного прогресса в решении всех трёх задач. Накоплены знания в различных областях, включая обработку информации, созданы мехатронные и информационные технологии, успешно расширяется применение конвейерных автоматизированных и роботизированных производств – всё это создаёт благоприятные условия для создания СИИ.

Проблемным вопросом остаётся создание теории СИИ. Если ремёсла и технологии развивались на основе практических действий, то представления о способах управления этими процессами за счёт активности НС строились чисто умозрительно, на основе субъективных ощущений, и на протяжении тысячелетий были весьма наивными. В последние 100-150 лет начали появляться научные теории о различных аспектах работы НС, а с конца 1950-х гг. – и практические результаты в решении «интеллектуальных» задач на вычислительной технике (хотя Лейбниц ещё в XVII в. рассматривал свой арифмометр как прообраз «интеллектуального» устройства). Но, по-

сколькo пока нет единого, общепринятого понимания, что такое «сознание», вопрос, можно ли действительно создать СИИ или только продолжать «имитировать» интеллектуальную деятельность, остаётся открытым.

5. Информация и её носители

Прекрасные идеи высказываются Д.И. Дубровским: «информация не может существовать без своего носителя» [6]. А информация – это то, что описывает *субъективную реальность* (СР), которая является неотъемлемой частью сознания в наших головах. При этом пусть информация нематериальна, но её носитель всегда материален и может быть исследован научными методами. Сам факт, что этим вполне понятным идеям Д.И. Дубровский посвятил значительное число работ, показывает, что для философов возможность отображения СР на нейросетевых структурах совсем не очевидна.

Сам Д.И. Дубровский пишет [6]: «Специфика явлений СР состоит в том, что им нельзя приписывать физические свойства (массу, энергию, пространственные характеристики). Они претендуют на особый онтологический статус, определение которого всегда предъявляло трудные вопросы для философов и естествоиспытателей, для тех, кто изучал связь психических явлений с деятельностью головного мозга».

Но и информации физические свойства (массу, энергию, пространственные характеристики) тоже приписывать нельзя. Более того, все компьютеры работают с абстрактными (не имеющими физической размерности) величинами, а смысл вычислениям придаёт написанная человеком программа, использующая численные результаты измерения веса, длины и других (размерных) физических величин в «правильных» формулах.

Аналогично в НС животных и человека сенсорные данные различной природы преобразуются в «безразмерные» импульсы активности нейронов, которые по аксонам передаются для обработки в «правильные» отделы мозга. А «программа» формирования «правильной» структуры связей записана в коде ДНК, который был сформирован в процессе эволюции путём естественного отбора – выбраковки неудачных вариантов «программ».

Приписывать физические свойства и проводить измерения можно носителям информации – элементам НС, но, в связи со сложностью НС, вопрос всё равно остаётся «трудным».

6. Необходимость конструктивных теоретических представлений

С последним, со сложностью строения НС, практически все согласны. Важны не только, и даже не столько опытные знания о строении и свойствах НС человека и животных, сколько теоретические взгляды на основные принципы её (и СИИ) построения. Без теоретических представлений объединить разнородные и зачастую противоречивые знания о НС высших животных и

человека вряд ли удастся. При этом теории не должны ограничиваться общими замечаниями о материальности (или нет...) сознания, а быть конструктивными, пригодными для разработки алгоритмов (желательно самоорганизующихся в связи со сложностью описываемых ими процессов) преобразования данных в НС и, в перспективе – СИИ. И создание таких теорий – центральная проблема для разрешения задач на пути создания СИИ

Современные нейросетевые успехи ИИ могут создавать впечатление, что все основные проблемы на пути к созданию СИИ решены. Хотя даже «крёстные отцы» глубокого обучения (на основе backpropagation) явно указывают на имеющиеся недостатки современных нейросетевых моделей [8], многие «продвинутые» сторонники нейросетевых решений настаивают, что кроме глубокого обучения (Deep Learning, DL) и обучения с подкреплением (Reinforcement Learning, RL), больше ничего и не надо [8].

Как справедливо указывает К. Фристон [9], проблема создания СИИ не сводится только к оптимизации, которую, действительно, современные нейросети решают достаточно успешно, нужны ещё методы поиска новых знаний. Оптимизация тоже является формой обучения, получения новых знаний, но только для удачно сформулированных задач, для которых достаточно данных для набора достоверной статистики и вычислительных мощностей для их обработки. Предлагаемый Фристом байесовский подход, к сожалению, не решает данной проблемы, поскольку основан на использовании параметров выделенных человеком простых объектов сложного мира. Когда задача грамотно поставлена, определены параметры, то, конечно, её можно решить и байесовскими, и рядом других методов. Но СИИ должен не только уметь решать удачно заданные человеком задачи, а ставить их самостоятельно, на основе анализа окружающей среды.

7. Сложность мира и возможности декомпозиции

Состояния сложной среды никогда не повторяются, «в одну реку нельзя войти дважды». Это является препятствием для набора достоверной статистики о её состояниях. Но если среда состоит из простых объектов и явлений, то для них можно выявить статистически достоверные свойства. Именно этим простые объекты и явления отличаются от сложной среды в целом – возможностью их описания на основе статистики данных, полученных в процессе наблюдения.

Представляется, что центральной проблемой на пути к созданию СИИ остаётся разработка и алгоритмическая реализация методов декомпозиции, применимых к широкому классу задач. *Декомпозицией* называется общий приём, состоящий в разделении проблемы на множество частных задач, с помощью объединения решений которых можно сформировать решение исходной проблемы в целом. Принято указывать, что в литературе первое упоминание о декомпозиции содержится в философском трактате Рене Де-

карта [11], который считается переломной работой, ознаменовавшей начало эпохи философии Нового времени и современное научное познание.

Поскольку методами декомпозиции ещё в доисторические времена пользовались и люди, и животные при достижении своих целей, нельзя сказать, что именно Декарт изобрёл их. Просто он одним из первых в явном виде описал их как важную часть процесса «правильного» мышления, но, вероятно, внимательное прочтение древних текстов может дать и более ранние упоминания об этом свойстве мышления. Но и до, и длительное время после публикации Декарта [10] в 1637 г. в Лейдене, декомпозиция рассматривалась как свойство мышления, получаемое «от Бога», без попыток описать его алгоритмическую природу.

В XX в. ситуация изменилась. Успехи теории алгоритмов, развитие программирования позволили создать значительное число алгоритмических подходов к проблеме декомпозиции, ряд которых описан, например, в [11]. В XXI в. методы декомпозиции получили новый импульс в связи с развитием средств параллельных вычислений [12]. Но общим недостатком всех используемых методов являются их узкая направленность на решение специальных задач и, главное, их применимость к задачам, уже поставленным человеком, а не к процессу постановки задачи.

8. Математические методы улучшения представления знаний

Тем не менее, сам факт наличия множества алгоритмизированных методов декомпозиции позволяет отказаться от мысли о «божественной» природе данного метода мышления и развивать новые, более универсальные методы, применимые к процессу постановки задач. При этом не следует думать, что создание таких методов позволит сразу решить все проблемы. Необходимо при создании СИИ использовать не только идеи оптимизации и декомпозиции, но и другие эффективные математические идеи.

Например, *увеличение степени локализации* представления знаний в структуре нейросетей будет не только лучше соответствовать идеям декомпозиции, но и позволит оптимизировать отдельные области знаний, не тревожась о сохранности сведений, не относящихся к данной области. *Линеаризация* представления выявленных нелинейных зависимостей даст возможность существенно упростить решение задач аппроксимации и построения прогнозов.

Не следует думать, что в современных «глубоких» нейросетях, для которых декларируется распределённая структура памяти, декомпозиция, локализация и линеаризация представления знаний полностью отсутствуют. Ниже будет обсуждено, почему оптимизация на основе градиентного спуска может поддерживать реализацию как указанных (декомпозиция, локализация и линеаризация) идей улучшения представления знаний, так и других подходов, позволяющих улучшить оптимизируемую суммарную оценку выполняемых преобразований. Проблема состоит не в начале ис-

пользования серьёзных математических идей, а в обеспечении более высокой эффективности их применения, переходу от вероятностного появления в структуре сети декомпозиции, локализации, линеаризации и пр. идей к направленному их использованию за счёт изначальной направленности заложенных в структуру нейросети алгоритмов их реализации.

Одним из важных плодов всё возрастающих современных успехов в решении «узких» задач ИИ на основе нейросетевых алгоритмов является смещение общественного мнения от «полного отрицания возможности воспроизводить интеллектуальную деятельность человека» к постепенному признанию наличия человеческих способностей у «интеллектуальных» машин. Студия VisionLabs с 2017 г. проводит международный ежегодный саммит Machines Can See [13]. За годы проведения саммита возможности машин по восприятию и интеллектуальной обработке сенсорной информации значительно увеличились, как количественно, так и качественно. И мы вправе в ближайшие годы ожидать проведения саммитов более продвинутых направлений (рис. 1).

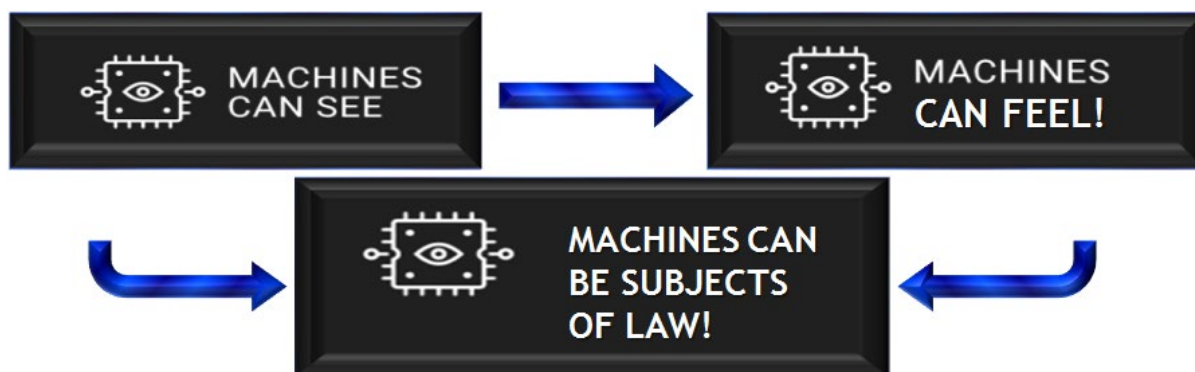


Рис. 1. Современные и перспективные возможности машин

9. Современные достижения

На сегодняшний день глубокие нейросети с обучением на основе метода обратного распространения ошибки остаются наиболее востребованными в задачах ИИ. Но если в 2010-е гг. было принято обсуждать архитектуру и алгоритмы работы нейросетей, то сейчас центральной является тема потребительских свойств программ и сетевых сервисов, работающих на нейросетевой основе. Например, в [14,15] можно посмотреть популярные (условно)бесплатные приложения, позволяющие обрабатывать и генерировать тексты, изображения, устную речь и переводить одно в другое в разных комбинациях, а также менять стиль изображений и язык, интонацию и голос, произносящий тексты (рис. 2).

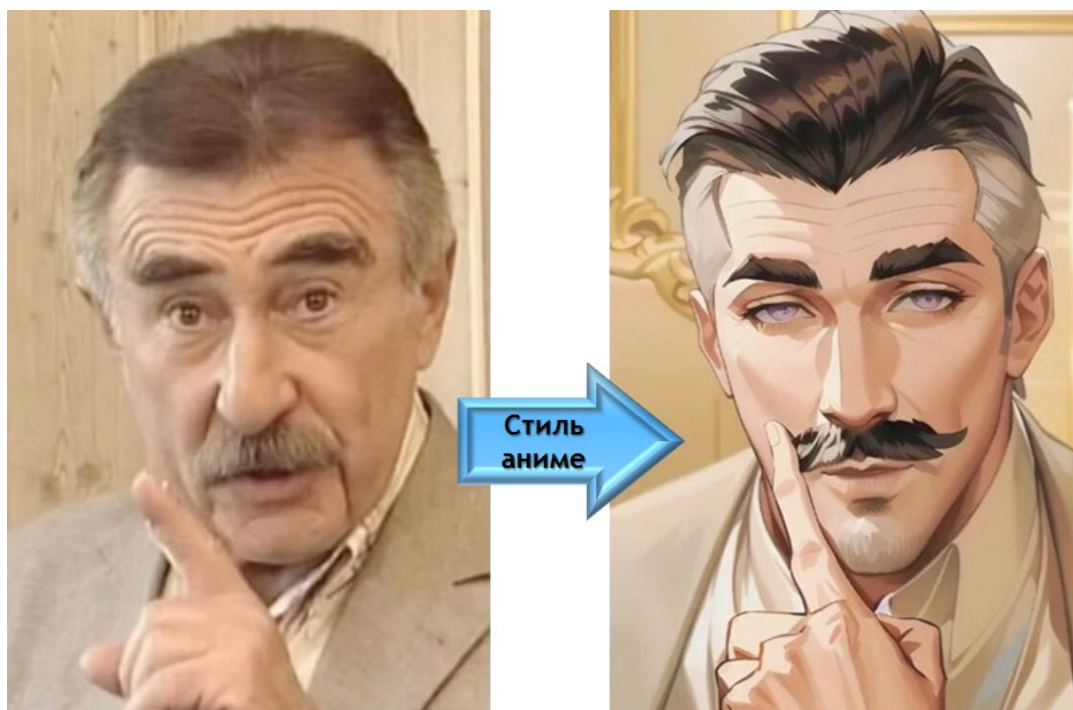


Рис. 2. – Пример стилизации портрета нейросетью

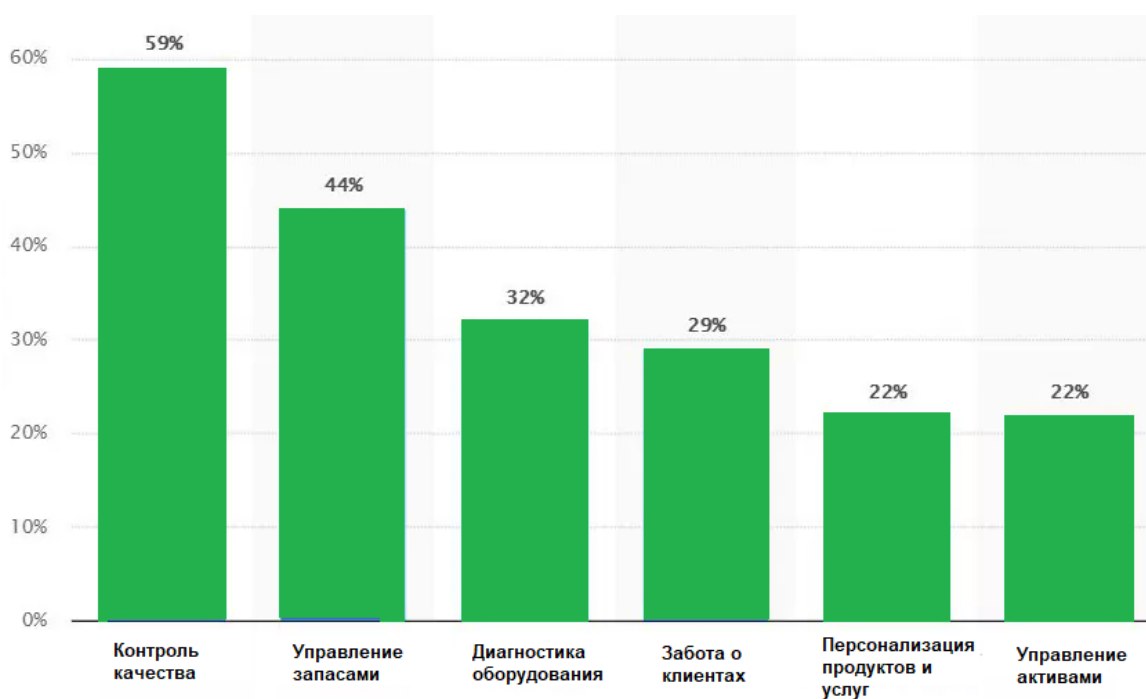


Рис. 3. Процент промышленных предприятий Западной Европы и США, применяющих нейросети в различных областях своей деятельности в 2022 г. [16]

И это только свободно доступные нейросетевые приложения. Нейросетевые ИИ-приложения находят очень широкое применение в самых разнообразных отраслях промышленности, в поисковиках, системах видеонаблюдения, автопилотах и многих других (рис. 3).

Важной особенностью современного этапа практического использования ИИ на основе нейросетей является отказ от обсуждения с (потенциальными) клиентами внутренней структуры нейросетевых алгоритмов. Сообщаются только потребительские свойства продуктов и инструкция по использованию, как сейчас принято для практически всех товаров.

Более того, разработчики конечных версий продуктов зачастую могут ограничиваться общими знаниями о свойствах используемых в них блоков. Специалисту по применению нейросетей, как правило, достаточно представлять, какие программно-аппаратные продукты можно использовать для решаемой ими задачи и знать, как их получить (скачать) с набирающих всё большую популярность открытых библиотек (рис. 4) и согласовать работу используемых компонентов.

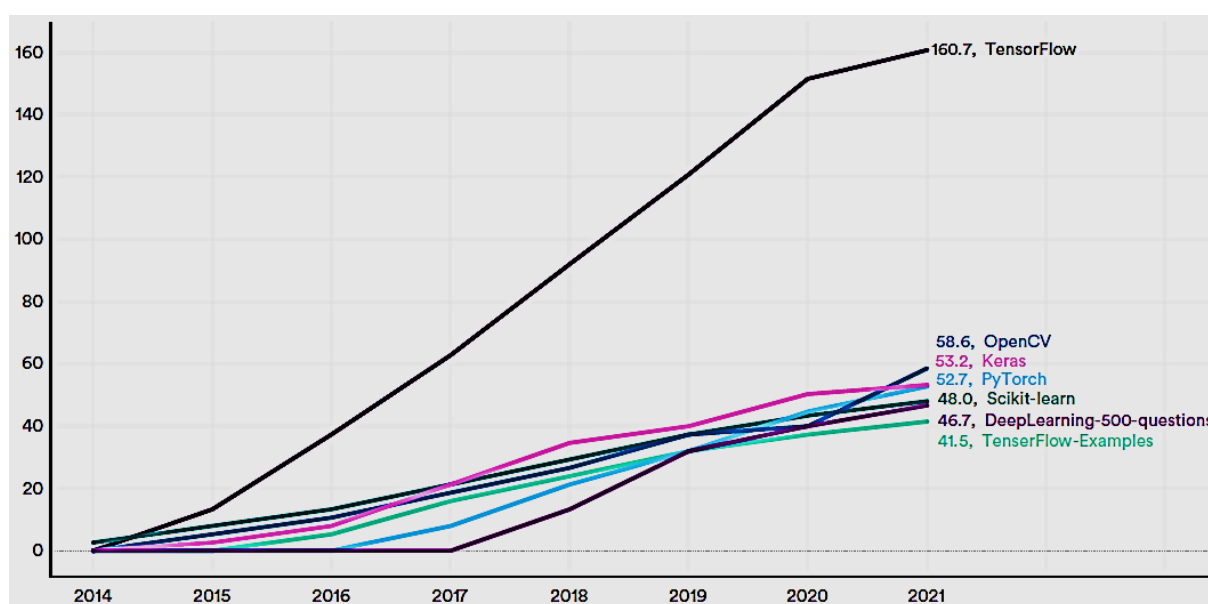


Рис. 4. Количество звёзд (отмеченных пользователями ссылок, в тыс.) GИTНUB от AI LIBRARY (для библиотек с более 40 тыс. звёзд в 2021 г.) [17]

Значительная часть бюджетов национальных программ по развитию ИИ используются для создания именно конечных, имеющих коммерческие приложения, продуктов. Ведущими мировыми центрами развития ИИ являются Китай, США и Западная Европа, предоставляющие своим исследователям и разработчикам в сумме более 2/3 мирового финансирования в данной области [18]. Относительный вклад не может быть точно измерен ни объёмами финансирования, ни количеством публикуемых журнальных статей. Тем не менее, некоторое представление об относительных позициях ведущих центров такие данные дают (см. рис. 5).

К сожалению, вклад России в мировой процесс развития ИИ невелик, он составляет порядка 0,1% от мировых объёмов финансирования и количества журнальных публикаций. Некоторым объяснением имеющейся статистики могло бы служить то, что в России больше уделяется внимания

теоретическим исследованиям, однако руководство наукой сейчас отдано в руки менеджеров и они стремятся переломить ситуацию на западный лад, отдавая предпочтение коммерческим разработкам и отменяя финансирование теоретических исследований.

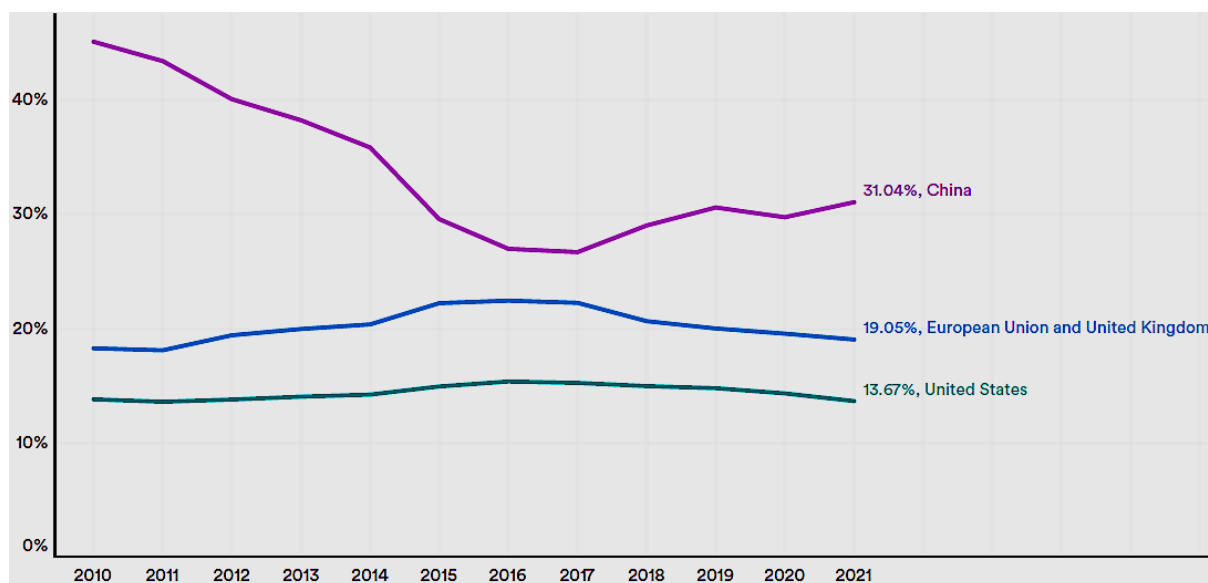


Рис. 5. Процент журнальных публикаций по регионам от общемирового количества статей по ИИ [17]

10. Теоретические достижения

Хотя разработки, имеющие близкие коммерческие перспективы обладают приоритетом при получении финансирования, теоретическим разработкам в мире тоже уделяется достаточное внимание. Ведущие мировые центры развития ИИ и крупнейшие корпорации создают и поддерживают работу исследовательских центров в области ИИ в целом и нейросетевых алгоритмов как современной основы прогресса в области ИИ.

Как и для коммерческих проектов, так и для теоретических разработок имеются приоритетные направления. Это развитие тех подходов и архитектур, которые уже продемонстрировали свою (коммерческую) успешность при решении практических задач.

Можно выделить три основные, наиболее популярные в этом смысле направления нейросетевых исследований. Это: 1) собственно глубокое обучение; 2) обучение с подкреплением; 3) генеративные сети.

1. *Глубокое обучение.* Идея и базовая теория «обратного распространения ошибки» (backpropagation error) была опубликована в [19] ещё в 1986 г. (и аналогичные идеи высказывались другими авторами ранее, в 1960-70-х гг.). Но, хотя попытки применения этой теории осуществлялись ещё в конце 1980-х, потребовалось порядка 25 лет, чтобы на их основе появились коммерчески успешные продукты.

И дело не столько в развитии высокопроизводительной вычислительной техники (НРС) и методов сбора данных (Интернет), сколько в необходимости для достижения успеха обучения использовать не голую идею градиентного спуска при настройке параметров нейросетей, а сочетать её с рядом алгоритмов, обеспечивающих устойчивость и достаточную скорость сходимости процесса обучения. Можно перечислять множество алгоритмов, направленных на начальную настройку весов связей и регулирование скорости обучения, но переход к действительно глубоким нейросетям, содержащим сотни и тысячи слоёв, произошёл только в 2015 г. после разработки метода батч-норм [20]. Идея нормализации активности слоёв, очевидно, обеспечивает устойчивость распространения сигналов (без затухания и чрезмерного «разгорания») по нейросети. Но при её практической реализации есть несколько тонких моментов, освоить которые до 2015 г. не удавалось, и нейросети содержали не более 10-20 слоёв.

То, что нейросети из сотен и тысяч слоёв работают лучше, чем те, которые содержат единицы или десятки слоёв – экспериментальный факт. Но причины превосходства «глубоких» нейросетей до сих пор остаются предметом обсуждения. Мы считаем, что с возрастанием числа слоёв увеличивается вероятность реализации более серьёзных идей представления знаний, чем простая оптимизация. Если при взаимодействии двух слоёв вероятность поддержки формирования таких способов улучшения представления знаний, как локализация, декомпозиция и линеаризация исчезающе мала, то будучи умноженной на 1000 слоёв она становится близкой к единице.

Как минимум, для локализации и декомпозиции в современных архитектурах нейросетей созданы условия. Свёрточные нейросетти производят локальный анализ сигналов, то есть осуществляют их декомпозицию на локальные компоненты. В рекуррентных нейросетях используются специальные компоненты, выполняющие функцию ворот (*gates*), открывающие и закрывающие пути записи и считывания различных данных в зависимости от сложившейся ситуации. Это позволяет сохранять информацию об относительно далёких по времени участках сигнала и осуществляет разбиение сигнала на отдельные компоненты. Но это не значит, что не надо пытаться повысить эффективность использования математических способов улучшения представления знаний, сделать их реализуемыми не вероятностно, как сейчас, а за счёт детерминированно реализуемых алгоритмов.

2. *Обучение с подкреплением.* Его отличие от обучения «с учителем» состоит в том, что подкрепление не указывает желательное направление изменения осуществляемого преобразования, а только оценивает по критерию «лучше–хуже». Системы и устройства, способные к обучению с подкреплением, (относительно) самостоятельно формируют вид преобразований и последовательности выполняемых действий, что значительно лучше соответствует идеи СИИ, чем обучение с учителем.

Конечно, сигнал подкрепления может выдавать и учитель, но в данном случае для его замены проще сформировать скалярный критерий оценки успешности, чем векторные образы оптимальных действий. Хотя игра в «холодно–горячо» не представляется трудной, в сложной среде (с неповторяющимися состояниями) и в условиях, когда сигнал подкрепления подаётся редко (в пределе – один раз, в конце последовательности действий), необходимо решать ряд проблем. Они связаны с необходимостью аппроксимации оценок для состояний, которые не получали сигнала подкрепления и, что хуже, ранее никогда не встречались.

Тем не менее, способы обучения марковских сетей были распространены на случай использования нейросетей и были разработаны методы, такие как DQN (deep Q networks [21]) и другие, что позволило решить многие современные проблемы обучения с подкреплением на нейросетях.

3. *Генеративные сети.* Генерация (создание) новых сигналов только отчасти напоминает обучение с подкреплением, поскольку в обоих случаях необходимо построить выходной сигнал, вид которого заранее никому не известен. Но в данном классе задач есть своя особенность: даже критерий оценки «лучше–хуже» формализовать крайне сложно.

Наиболее распространённым путём решения этой проблемы является использование генеративных состязательных сетей (GANs, [22]). Сгенерированный сигнал не сравнивается с эталонным и не оценивается по некоторому критерию, а для оптимизации процесса генерации организуется соревновательный процесс между двумя нейросетями. Одна генерирует новые сигналы, а вторая учится отличать сгенерированные сигналы от реальных, которые вперемешку со сгенерированными предъявляются для классификации. Обучаются обе сети: первая в случае, если вторая правильно классифицировала сигнал как сгенерированный, а вторая – если она ошиблась при классификации сигнала. Если какая-то из сетей обучилась так, что при текущем уровне обучения другой сети перестаёт ошибаться в генерации или классификации, то это приводит к тому, что другая соревнующаяся сеть получает сигналы на изменение своих параметров – их оптимизацию. Это позволяет уменьшить число её ошибок, значит появятся сигналы на обучения её «напарницы».

Как и в глубоком обучении, и в обучении с подкреплением, в генеративных сетях есть много тонкостей, без которых они не работают эффективно. В некоторых случаях можно обойтись без соревновательного процесса. Но в целом, для генерации сигналов необходимо сформировать модель класса сигналов в скрытых – латентных – переменных (активностях элементов внутренних слоёв нейросети) и соотнести эту модель с подаваемыми на вход параметрами, задающими свойства генерируемого сигнала.

Формирование моделей классов генерируемых сигналов – это способ представления знаний о сложном окружающем мире. Все наиболее успешные и пользующиеся популярностью нейросети используют идеи генера-

ции сигналов. Это и самые разнообразные GANs, и «диффузионки», такие, как DALLE 2, Midjourney [23], Imagen, и различные версии систем на основе GPT (Generative Pre-trained Transformer) – GPT3 [24], ChatGPT, GPT4. Возможности генерации сигналов, реализованные путём формирования их моделей при условии обучения на больших объёмах широкомасштабных данных, производят сильное впечатление на пользователей, поскольку позволяют формировать содержательные ответы на самые разнообразные запросы. Такие большие системы ИИ, построенные на генеративных нейросетях сейчас принято называть «базовыми моделями».

Этот термин (a foundation model) популяризировал Центр исследований базовых моделей (Center for Research on Foundation Models, CRFM) Стэнфордского института человеко-ориентированного искусственного интеллекта (Human-Centered Artificial Intelligence's, HAI) [25]. Базовые модели заметно повлияли на способы построения систем ИИ с момента их появления в 2018 г. Для привлечения интереса к базовым моделям даже делаются заявления, что это и есть первые образцы СИИ [26]. Обоснованность таких заявлений вызывает споры, но то, что это важный шаг на пути построения СИИ, – многими признаётся.

11. Будущее СИИ

Современные успехи создания систем ИИ на базе нейросетевых подходов привели к формированию целой отрасли промышленности, в которой задействованы ведущие мировые исследовательские и учебные центры, крупные технологические и сетевые компании, налажено крупносерийное производство предназначенных для нейровычислений микросхем и устройств. И, главное, сформирован рынок объёмом в десятки млрд. долларов, на котором интеллектуальная продукция пользуется устойчивым спросом. Это даёт устойчивые финансовые гарантии для успешного развития направления на построение СИИ. И оптимистичный прогноз состоит в том, что уже в ближайшие 3-10 лет СИИ удастся построить.

Как нам представляется и уже указывалось в пп. 8-11, основными направлениями развития работ в области ИИ должно стать целенаправленное применение математически обоснованных методов повышения эффективности представления знаний, таких, как локализация, декомпозиция и линеаризация и других. Благоприятные условия достижения существенных результатов для обеспечения теоретического обоснования создания СИИ уже сложились, а технологическая и промышленная поддержка уже имеется. Так что до создания СИИ остаются если не месяцы, то единицы лет.

Естественно, что имеющиеся успехи в создании ИИ, не являющегося «сильным», не только послужат основой создания СИИ, но и будут продолжать развиваться и активно использоваться в тех задачах, для которых понимание, сознание, эмоции и прочие «человеческие» качества не нужны. Более того, ещё в течение довольно длительного времени (даже после со-

здания СИИ) устройства и системы ИИ будут определять уровень технологического развития как отдельных стран, так и цивилизации в целом.

12. Зачем нам СИИ?

А что такое СИИ? Определений этому понятию предлагается множество, каждый автор указывает на недостатки ранее данных определений и пытается дать своё. В целом, СИИ (artificial general intelligence, AGI) принято определять, как систему способную понимать или изучать любую интеллектуальную задачу, с которой могут справиться люди или другие животные [27]. При этом предполагается, что создание системы, обладающей интеллектуальными способностями на уровне человека, позволит легко нарастить уровень её возможностей далеко за пределы человеческих, как это происходит с большинством машин, созданных человеком. И эти «сверхчеловеческие» возможности создают опасность разрушения сложившихся общественных отношений, что приводит к появлению всевозможных «Движений за дружественный СИИ», ярким представителем которых является, например, «Институт будущего жизни» (Future of life institute, FLY [28]).

При этом такие движения радикально отличаются от луддитов тем, что тогда станки заменили простых ткачей, а создаваемый сейчас СИИ может потеснить интеллектуальную элиту. Поэтому «Движения за дружественный СИИ» организуют не рабочие и даже не простые программисты или учёные, а те, кто причисляет себя к интеллектуальной элите по причине обладания или приближённости к власти.

Они часто поднимают вопрос: а нужен ли человечеству СИИ? Может быть, не стоит выпускать джинна, давайте ограничимся развитием ИИ, который будет бесстрастно выполнять задачи поставленные человеком и не станет формировать собственные цели? И даже выдвигаются утверждения, что создавать СИИ, который сможет самостоятельно определять для себя задачи аморально, это положит конец господству человека на Земле и приведёт к гибели человечества.

13. СИИ – для консервации сложившегося миропорядка?

Когда такие заявления делаются такими богатыми и знаменитыми персонами, как Гейтс, Маск и Безос, или излагаются в книгах известных специалистов, вроде С. Рассела [29] и М. Тегмарка [30], то их можно понять. Находясь в положении обладающих или обласканных властью персон, они не стремятся изменить свой статус. С их точки зрения будет лучше, если они так и будут продолжать править миром. Обосновывают свою позицию они тем, что вот сейчас же они правят миром – и всем хорошо, а если перестанут, и их заменят не такие же, как они, а люди с другими целями и принципами или, что ещё опаснее, СИИ, то это может привести к большим бедам, катастрофам и даже гибели человечества.

Именно поэтому С. Рассел в книге [31] прямо пишет, что следует создать и совершенствовать ИИ в том направлении, чтобы он умел угадывать желания таких персон, как он сам, и его спонсоров. Хотя Рассел в своей книге справедливо отмечает, что желания людей, особенно богатых и властных, как правило, ужасны. И во многих сказках, где джинн выполняет три желания, третье состоит в отмене первых двух, уже исполненных желаний. Но это и ряд других противоречий, которые пока не разрешены, не останавливают Рассела от вывода, что нужен только такой ИИ, который сможет угадывать и исполнять желания (состоятельных) людей.

И это одно из наиболее вероятных направлений развития ИИ. Поскольку тех. задания и гранты на исследования выдают персоны, обладающие властью, то следует согласиться со С. Расселом, он вполне успешно выразил их позицию по вопросу развития ИИ.

Но не все считают, что они живут так же или почти так же хорошо, как Гейтс, Маск и Безос. Более того, многие хотят занять их место и рассматривают создание СИИ как важный инструмент в борьбе за власть и богатство. Они отдают себе отчёт, что если их СИИ будет слабее, чем у конкурентов, то в борьбе за мировое лидерство победить будет не просто сложно, а практически невозможно.

Так, в своей книге «Сверхдержавы ИИ» [32] известный специалист в данной области, К.-Ф. Ли пишет, что когда программа АльфаГо в марте 2016 г. выиграла матч со счётом 4:1 у Ли Седоля (одного из 4-х лучших в мире игроков в го), то это произвело на коммунистическое руководство Китая даже большее впечатление, чем на США – запущенный СССР в 1957 г. первый спутник Земли. Китайские власти имели основания рассматривать свою страну как наследницу древней цивилизации, строящую самое прогрессивное общество. Появление у политического конкурента – США – интеллектуальной программы, способной выигрывать в национальную китайскую, пусть настольную, но стратегическую игру, подрывало декларируемую уверенность КПК в интеллектуальном превосходстве Китая.

В результате 8 июля 2017 г. Госсоветом Китая был выпущен «План развития искусственного интеллекта нового поколения» [32]. Даже в открытой части документа указывается, что «Искусственный интеллект стал новым направлением международной конкуренции. Искусственный интеллект — это стратегическая технология, за которой будущее. Крупнейшие развитые страны мира рассматривают развитие искусственного интеллекта как основную стратегию повышения национальной конкурентоспособности и поддержания национальной безопасности. Мы должны поставить развитие искусственного интеллекта на национальный стратегический уровень, твердо ухватиться за стратегическую инициативу международной конкуренции в новый этап развития искусственного интеллекта, создать новые конкурентные преимущества, открыть новое пространство для развития и эффективно защитить свою национальную безопасность».

Естественно, что руководство Китая не рассматривает свою программу развития ИИ как способ постепенного отказа КПК от власти. Наоборот, речь идет о создании в Китае на базе технологий ИИ новой системы для поддержания в стране социально-политической стабильности («системы социального кредита»). На всей территории КНР должна быть сформирована универсальная система оценки граждан, компаний и НКО. С помощью ИИ и в соответствии с алгоритмами, утвержденными государством, будут проанализированы все доступные данные о человеке или компании, включая их онлайн-поведение. Обладатели высокого рейтинга получают социальные и экономические льготы, обладатели низкого – подвергнутся ограничениям.

Таким образом, и США, и КНР, несмотря на все различия их социальных институтов, рассматривают развитие ИИ (и в перспективе – СИИ) как средство сохранения своей власти. Но это, согласно К.-Ф. Ли – сверхдержавы искусственного интеллекта. Если США предлагает создать однополярный мир (под руководством США), то Китай пока говорит о биполярном мире (США и Китай).

14. Многополярный мир как основа прогресса СИИ и цивилизации

Но не все страны готовы вернуться к временам холодной войны, в которой Китай заменит СССР в качестве противовеса США. Как заявил первый заместитель главы Еврокомиссии (ЕК) Ф. Тиммерман [33], «Мы живем не в биполярном мире, хотя эти две страны склонны полагать, что он биполярен. <...> Обратите внимание на такие государства, как Канада, Австралия и Новая Зеландия, они хотят видеть многополярное мировое сообщество». Далее он добавил, что, по его мнению, КНР и США в итоге поймут, что им необходимо существовать в условиях непростого многополярного мира.

Но для достижения такого понимания необходимо, чтобы ведущие (сверх)державы отказались в развитии ИИ и глобального управления процессом создания СИИ, в смысле ограничения его способностей, строить собственные, не продиктованные властными верхушками планы. В то же время другие страны, например, Япония, Израиль или Южная Корея, смогли бы добиться прорывных успехов в создании СИИ, которые дали бы им возможность выйти из-под контроля США.

Важно, чтобы при создании СИИ правила его развития определялись бы не из единого центра, а сохранялась конкуренция в развитии как СИИ, так и общественно-политических систем в целом. Только на основе соревнования идей и способов их реализации возможно продолжение прогресса цивилизации. Это, конечно, не значит, что конкуренция должна быть неуправляемой, осуществляться через глобальные конфликты и катастрофы. Наоборот, необходимы международные договоры, направляющие конкуренцию в цивилизованное, по возможности бесконфликтное русло. При

этом бесконфликтность должна обеспечиваться не консервацией лидирующего положения современных ведущих (сверх)держав, а созданием условий для сравнения и распространения лучших практик построения как СИИ, так и общественно-политических систем в целом.

15. Знания как основа развития самоорганизации

В случае, если удастся сохранить многополярный мир и направить развитие СИИ не на пользу властной верхушки однополярного мира, а на значительно более успешное решение задач общественного и технического прогресса всего человечества, то это приведёт не к «технологической сингулярности» [34], а к резкому, но конечному увеличению скорости научного и технологического развития цивилизации.

Процессы самоорганизации являются основой развития жизни и цивилизации и базируются на получении и накоплении новых знаний. Роль знаний в процессе функционирования автономных агентов ИИ (как живых организмов, так и автономных устройств) показана на схеме на рис. 6.



Рис. 6. Использование знаний при взаимодействии агентов ИИ со средой

Автономные агенты ИИ взаимодействуют со средой, получая данные через сенсоры, и выполняют действия (не обязательно механические) путём управления через эффекторы своим физическим (или виртуальным) телом.

Знания и входные данные используются при формировании законов управления эффекторами. Входные данные и процесс иерархического моделирования также используются для получения новых знаний.

В «узком» ИИ, как правило, процессы накопления и использования знаний бывают разделены и требуют внешнего управления. Важным отличием «узкого» ИИ от СИИ принято указывать ограниченность области знаний, которые может накапливать и использовать агент ИИ. И сейчас высказываются мнения, что базовым моделям (a foundation model) удалось это ограничение преодолеть – они способны обрабатывать данные для широкого круга областей знаний.

Мы не думаем, что базовые модели – это и есть СИИ. Дело в том, что их успех основан на анализе текстов, в которые заложены понятия, выделенные человечеством из сложного мира и описания соотношений между этими понятиями. Иными словами, это результат статистической обработки знаний, полученных человеческой цивилизацией за всю историю её развития. Это важное достижение, имеющее большую практическую пользу. Но способность базовых моделей писать тексты на, казалось бы, произвольные темы не даст им возможности на основе анализа текстов до XIX в. вывести законы квантовой физики. Обладание знаниями человечества важно. Оно позволяет не только писать тексты, но и ранжировать изображения или звуки, но процесс выявления новых знаний из непрерывной среды базовые модели пока не реализуют. То есть, превосходя человека в обширности доступных для использования знаний, они уступают в некоторых возможностях получения новых знаний.

Новые знания на основе моделирования игры в го и другие настольные игры (шахматы, сёги – японские шашки) ещё в 2017 г. научили получать новую версию программы АльфаГо – программу АльфаZero [35]. Если АльфаГо учили на записях партий гроссмейстеров, то АльфаZero получила навыки эффективной тактики в играх, имея нулевые начальные знания, просто играя со своей копией и постепенно улучшая свою тактику. Чтобы превзойти свою предшественницу АльфаГо (лучшую в мире программу игры в го), АльфаZero потребовалось 8 часов обучения. На достижение уверенного выигрыша у аналогичных программ в шахматы и сёги потребовалось 4 и 2 часа соответственно.

Но и АльфаZero, хотя и выиграла у ведущих мировых мастеров го со счётом 100:0, не была признана СИИ. Дело в том, что удалось продемонстрировать только превосходство нейросетевого подхода над эвристическими методами программирования. А выделять знания из сложного непрерывного мира и формировать их иерархическое описание без использования человеческих знаний тогда не потребовалось. За прошедшее с 2017 г. время команда Д. Хассабиса (создавшая АльфаГо и АльфаZero) добилась ряда других значительных успехов в развитии ИИ и, если их творчество не ограничат, то их прямое участие в создании СИИ весьма вероятно.

Выделение знаний из сложного непрерывного мира и формирование их универсального иерархического описания пока остаётся сложной задачей

для разработчиков ИИ. Но есть все основания считать, что в ближайшие годы она будет решена и именно на основе нейросетевых алгоритмов. Теоретическим обоснованием этого утверждения можно считать теорему Эшби [36] о необходимой сложности управляющей системы. И именно нейросетевые алгоритмы обладают наибольшим числом обучаемых параметров и потенциально способны строить наиболее сложные законы управления.

Конечно, никакое техническое устройство не сможет превзойти весь мир по своей сложности. Но наибольшим потенциалом строить управление в наблюдаемой локальной части мира (рис. 6) обладают нейросетевые алгоритмы, способные к самообучению

16. Выводы

Вопрос построения СИИ перестал быть областью обсуждения футурологов, а перешёл в плоскость практической реализации. В книге [37] проводится обзор тенденций движения в сторону СИИ. Но следует заметить, что дело не сводится к постепенному исключению человека из решения всё большего числа задач. Технологии ИИ, а тем более СИИ могут послужить значительному ускорению научно-технического прогресса. Последний в настоящее время является основой процветания и военно-политического могущества ведущих (сверх)держав.

При сохранении многополярного мира, цивилизованной конкуренции между странами, прогресс невозможно остановить. В ближайшие годы будут найдены технические пути выделять знания из сложного непрерывного мира и ставить на их основе новые задачи. После этого часть управленческих решений, связанных с постановкой задач, можно будет передать СИИ. Это значительно повысит качество управленческих решений, которые люди принимают далеко не всегда на основе взвешенного учёта всех доступных данных и часто в своих корыстных, не связанных с целями прогресса, интересах. Страны, которые смогут использовать технологии СИИ при формировании своей внешней и внутренней политики, получат преимущества в международной конкуренции (если она сохранится).

России, с её приверженностью к многополярному миру, необходимо уделять больше внимания развитию СИИ. И дело не только в обеспечении финансирования работ по ИИ на мировом уровне (что требует увеличения финансирования в данной области в 30-100 раз), но и в формировании цивилизованной конкурентной среды для создателей и разработчиков ИИ, а главное – поддержке научных разработок в области СИИ.

Литература

1. Хопкрофт Дж., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений. – М.: Вильямс, 2002. – 528 с.

2. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
3. McCulloch W.S., Pitts W. A logical Calculus of Ideas Immanent in Nervous Activity – Bull. Mathematical Biophysics, 1943.
4. Widrow B., Hoff M.E. Adaptive switching circuits. 1960 IRE WESTCON Conferencion Record. – New York, 1960.
5. Маккарти Дж, Мински М., Рочестер Н., Шеннон К.Е. Предложения для Дартмутского исследовательского проекта по искусственному интеллекту. <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>
6. Дубровский Д.И. [Проблема «Сознание и мозг» теоретическое решение.](#) – М.: «Канон+» РООИ «Реабилитация», 2015. – 208 с.
7. Bengio Y., Lecun Y., Hinton G. Deep Learning for AI // [Communications of the ACM. July 2021. 64\(7\), 58-65.](#)
8. Dickson B. DeepMind scientists: Reinforcement learning is enough for general AI. <https://bdtechtalks.com/2021/06/07/deepmind-artificial-intelligence-reward-maximization/>
9. Friston K. The free-energy principle: A unified brain theory? // [Nat. Rev. Neurosci. 2010, 11, 127-138.](#)
10. Декарт Р. Рассуждение о методе // Сочинения в 2 томах. – М., 1989–1994. Т.1.
11. Цурков В.И. Декомпозиция в задачах большой размерности / Под ред. Г.С. Поспелова. – М.: «Наука», 1981. – 352 с.
12. Васильевский Ю.В., Ольшанский М.А. Краткий курс по многосеточным методам и методам декомпозиции области. – М.: МГУ. 2007. – 103 с.
13. Саммит “Machines can see”. <https://visionlabs.ru/ru/company/events>.
14. Нейросети 2022 года, которыми может пользоваться каждый <https://thecode.media/neuro-2022/>
15. Топ 40 лучших нейросетей для каждого. <https://vc.ru/u/1216053-nftishka/605828-top-40-luchshih-neyrosetey-dlya-kazhdogo-luchshie-neyroseti-dlya-zarabotka>
16. Статистика применения нейронных сетей в промышленности. <https://upperator.ru/industry#rec432138272>
17. Artificial Intelligence Index Report. Stanford University, 2022.
18. Альманах «Искусственный интеллект». Индекс 2020. Аналитический сборник, М., №8, 2021.
19. Rumelhart D.E., Hinton G.E., Williams R.J. Learning Internal Representations by Error Propagation // Parallel Distributed Processing. Vol.1, pp. 318-362. Cambridge, MA, MIT Press, 1986.
20. Ioffe S., Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift // Proceedings of the 32nd International Conference on International Conference on Machine Learning. 2015. Vol. 37, Pp.448-456.

21. *Sanz M.* Ausin Introduction to Reinforcement Learning. Part 3: Q-Learning with Neural Networks, Algorithm DQN. <https://markelsanz14.medium.com/introduction-to-reinforcement-learning-part-3-q-learning-with-neural-networks-algorithm-dqn-1e22ee928ecd>
22. *Goodfellow I., Pouget-Abadie J., et al.* Generative adversarial networks. [arXiv:1406.2661](https://arxiv.org/abs/1406.2661)
23. *Rose J.* Inside midjourney, the generative art ai that rivals DALL-E <https://www.vice.com/en/article/wxn5wn/inside-midjourney-the-generative-art-ai-that-rivals-dall-e> (2022).
24. *Brown T., Brown T., et al.* Language models are few-shot learners. [arXiv:2005.14165](https://arxiv.org/abs/2005.14165)
25. Introducing the center for research on foundation models. <https://hai.stanford.edu/news/introducing-center-research-foundation-models-crfm>
26. *Vincent J.* ChatGPT proves AI is finally mainstream – and things are only going to get weirder. <https://www.theverge.com/2022/12/8/23499728/ai-capability-accessibility-chatgpt-stable-diffusion-commercialization>
27. https://en.wikipedia.org/wiki/Artificial_general_intelligence
28. Future of life institute, FLY. <https://futureoflife.org/>
29. *Russell S.* Human compatible: Artificial intelligence and the problem of control. – Viking, 2019.
30. *Tegmark M.* Life 3.0 : Being human in the age of artificial intelligence. – New York: Knopf. (2017).
31. *Лу К.-Ф.* Сверхдержавы искусственного интеллекта. – М.: «Манн, Иванов и Фербер», 2019. – 238 с.
32. План развития искусственного интеллекта нового поколения. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
33. В ЕК указали на заблуждения КНР и США о биполярном устройстве мира. <https://iz.ru/1435729/2022-12-05/v-ek-ukazali-na-zabluzhdeniia-knr-i-ssha-o-bipoliarnom-ustroistve-mira>
34. *Kurzweil R.* The singularity is near. – New York, NY: Penguin Group, 2005.
35. *Silver D., Hubert T., et al.* Mastering chess and shogi by self-play with a general reinforcement learning algorithm. [arXiv:1712.01815](https://arxiv.org/abs/1712.01815)
36. *Ashby W.R.* Principles of the self-organizing dynamic system // Journal of General Psychology. 1947. 37, 125-128.
37. *Ведяхин А. и др.* Сильный искусственный интеллект: На подступах к сверхразуму. – М.: Интеллектуальная Литература, 2021. – 232 с.